

# DSE4 : Final year Project



AI Chatbot : SAM

Amit Sarkar

021-1112-0453-21

0221CMSH-0146

# >



# Introduction



Ever have a question about Balurghat College but weren't sure where to look?

Now you can chat directly with our Chatbot. Chatbot Sam, is available 24/7 to help you navigate your college journey!



## Understanding Chatbot : Sam

Before diving into the creation process, it's crucial to understand the role and potential of this chatbot. This chatbot is not connected to the Internet. It can only answer college-related questions.



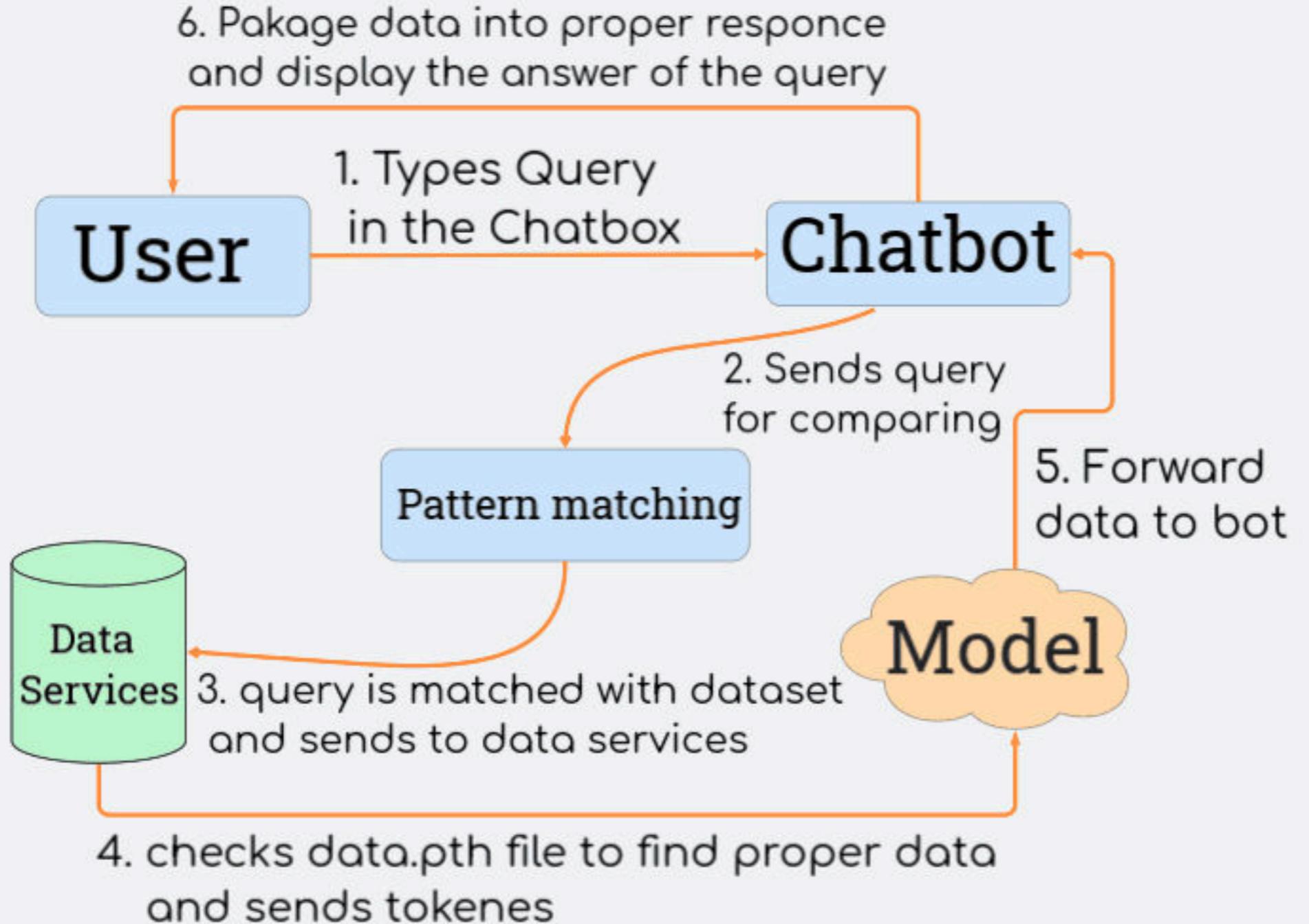
# Defining Objectives ← +

- >> Provide Immediate and Easy Access to Information
- >> Enhance the Student Experience

Overall, the chatbot aims to be a valuable tool for students by empowering them to find the information they need efficiently, ultimately creating a smoother and more positive college experience.

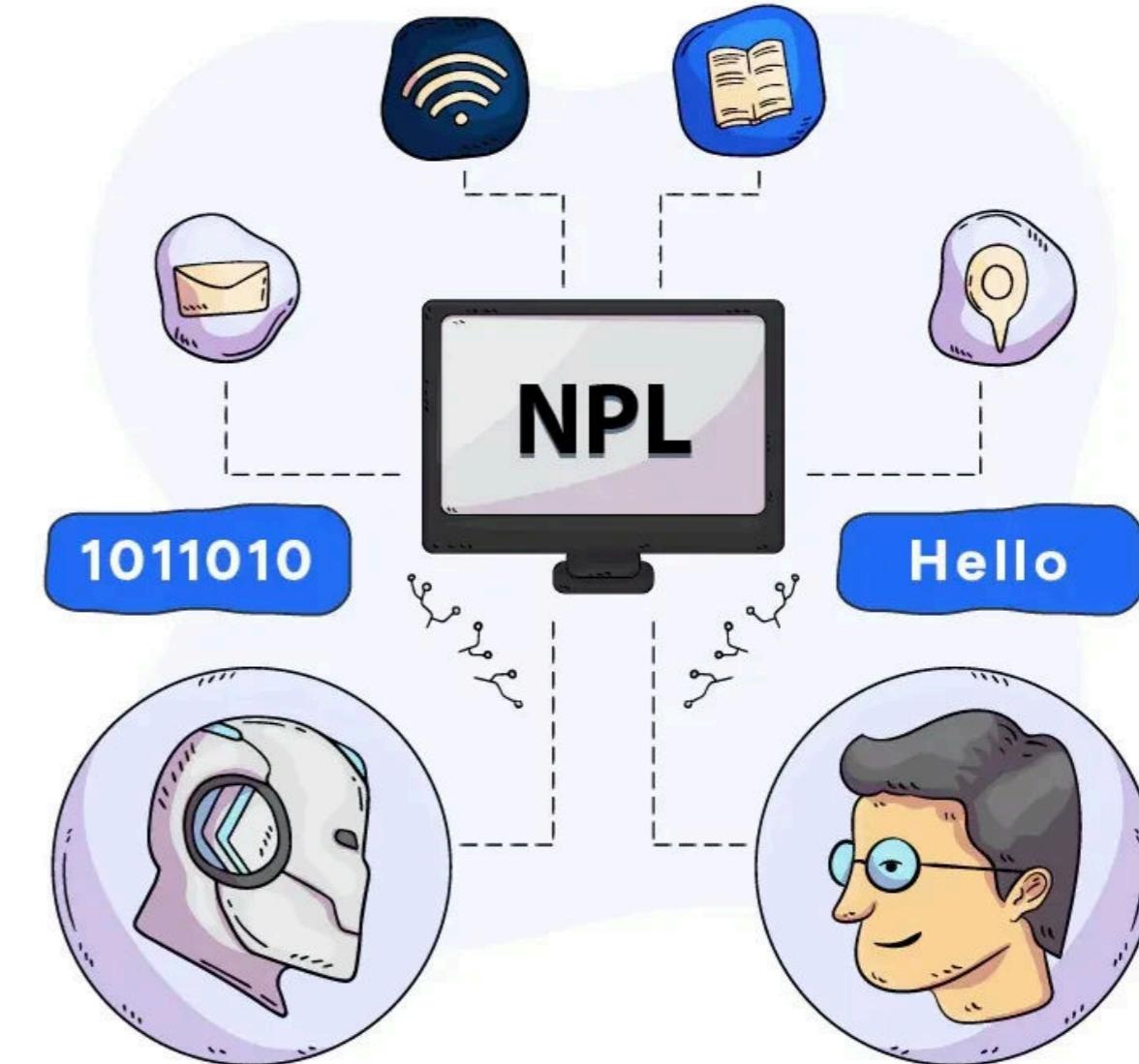


# Designing Conversational Flows



# Implementing Natural Language Processing

- Tokenization
- Stemming



- Tokenization

Tokenization , in Natural Language Processing (NLP) and machine learning, refers to the process of converting a sequence of text into smaller parts, known as tokens. These tokens can be as small as characters or as long as words.

→ "Hi , How are you ?

[ 'Hi' , 'How' , 'are' , 'you' , '?' ]

→ "What are you doing ?"

[ 'What' , 'are' , 'you' , 'doing' , '?' ]

- # Stemming

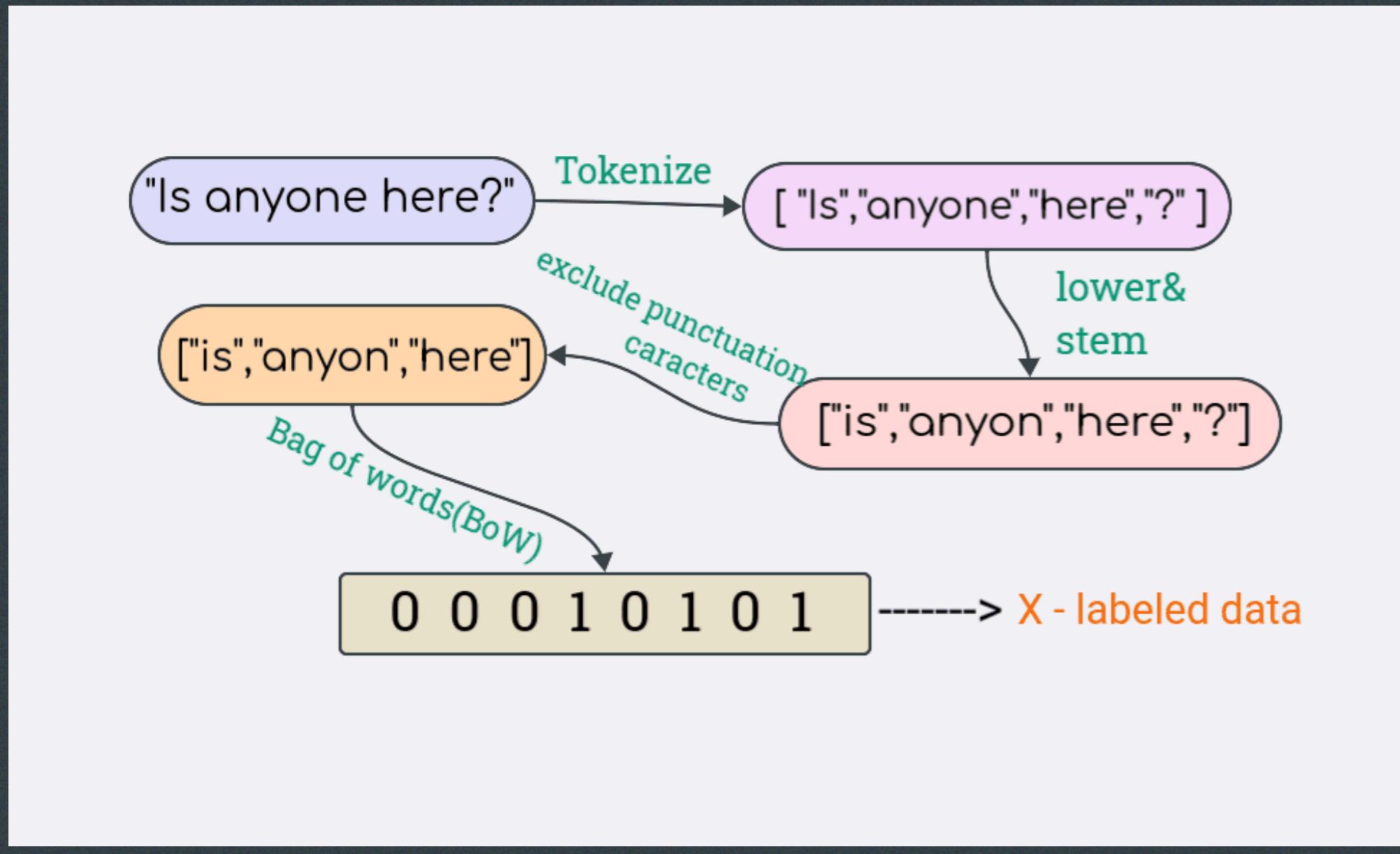
Stemming is a text preprocessing technique in natural language processing (NLP). Specifically, it is the process of reducing inflected form of a word to one so-called “stem,” or root form, also known as a “lemma” in linguistics.

[ "changing" , "changed" , "change" ]

-----

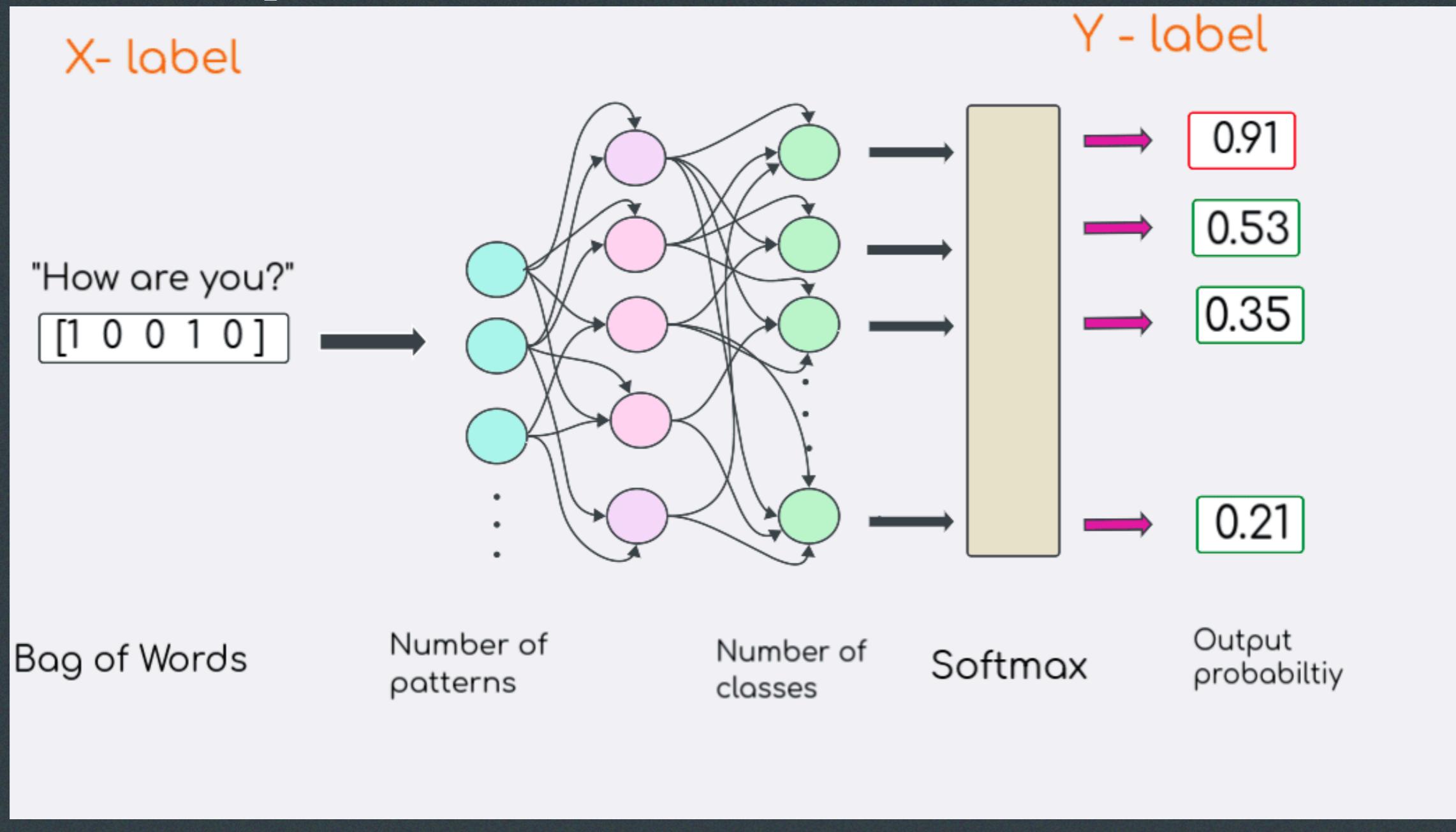
[ "chang" ]

# • NLP Processing Pipeline

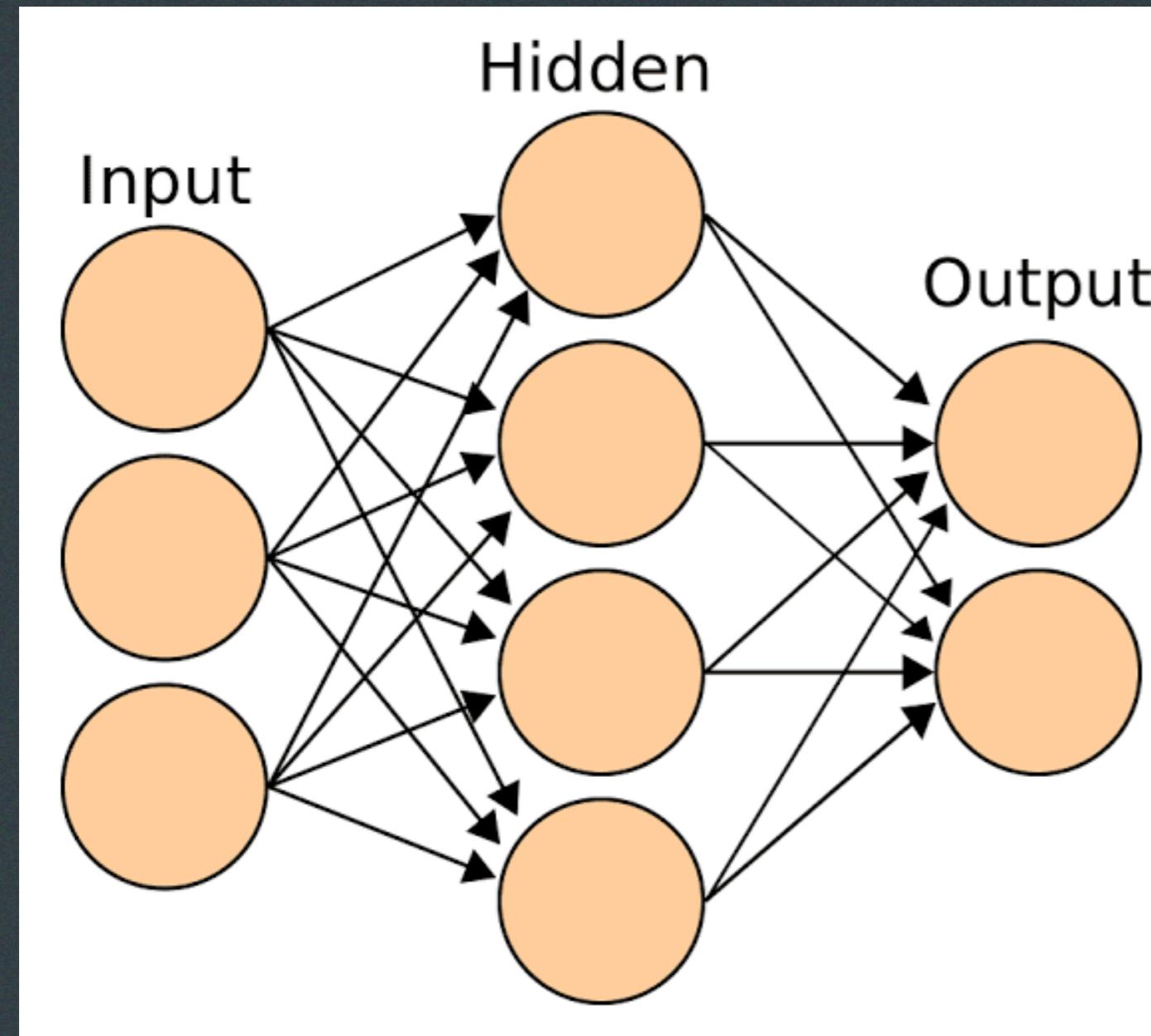


- It receives a string as input, tokenizes it to create tokens, lowerscases every character, afterwards it performs stemming to get the root word. It labels the data using the bag of words function after obtaining the root.

# • Implementation



# >> SoftMax Algorithm



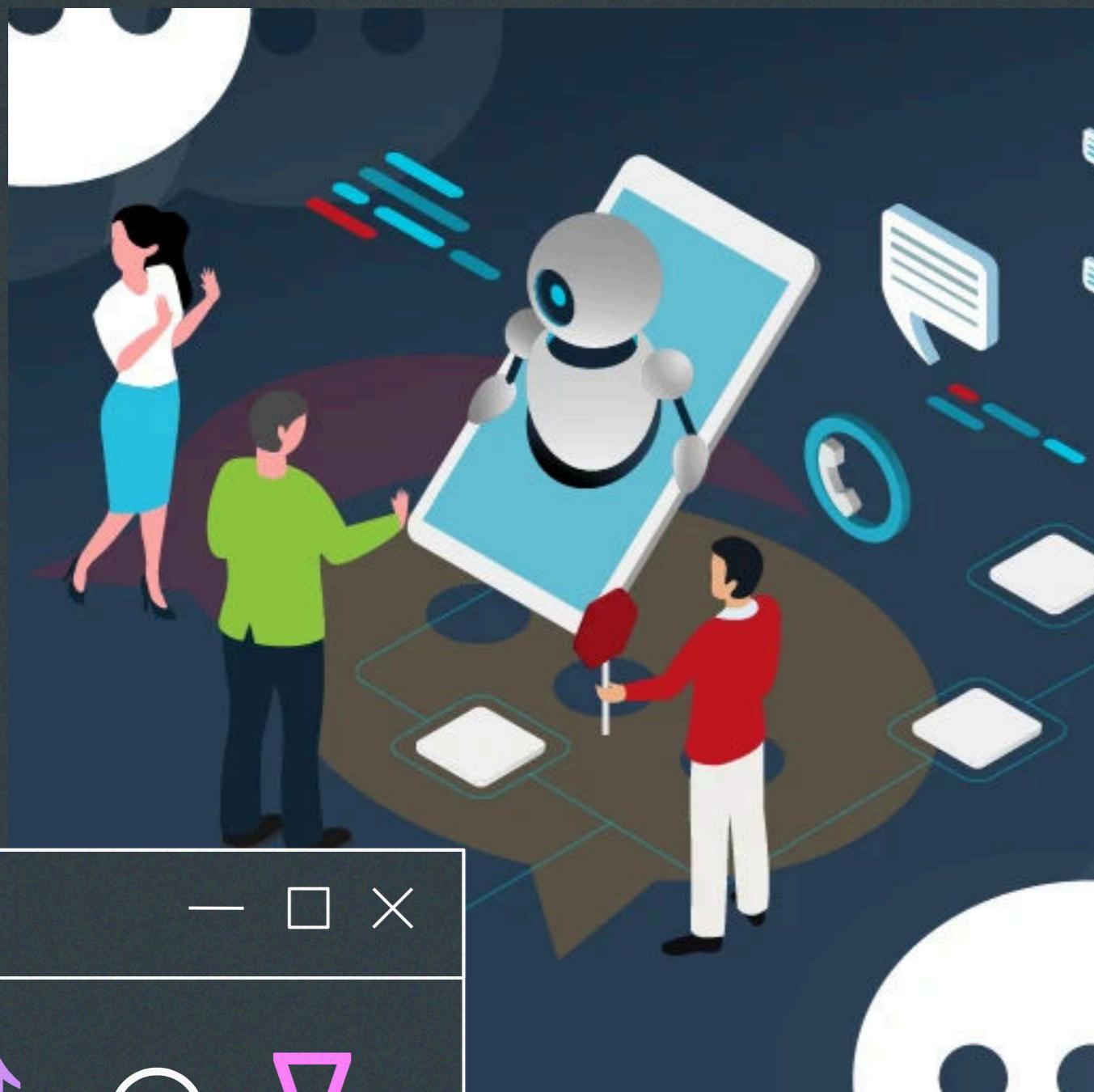
- The softmax function is an activation function in the final layer of a neural network.
- It is a multi-category equivalent of a sigmoid function and is used whenever there are more than two outcomes (e.g. non-binary). The probabilities of the categories must sum up to 1.

# Further Enhancement



- Multi-Lingual Support
  - We can add multi-Lingual support for students who prefer to interact in their native language
- Voice Chat Integration
  - We can integrate voice chat function that will enhance accessibility and convenience

# → Limitations



- Limited Scope

Sam will focus solely on college related information specific to Balurghat College. It won't be able to answer internet searches.

- Offline Dependence

Since Sam functions offline, its knowledge base may not reflect real-time updates or changes made on the college website.

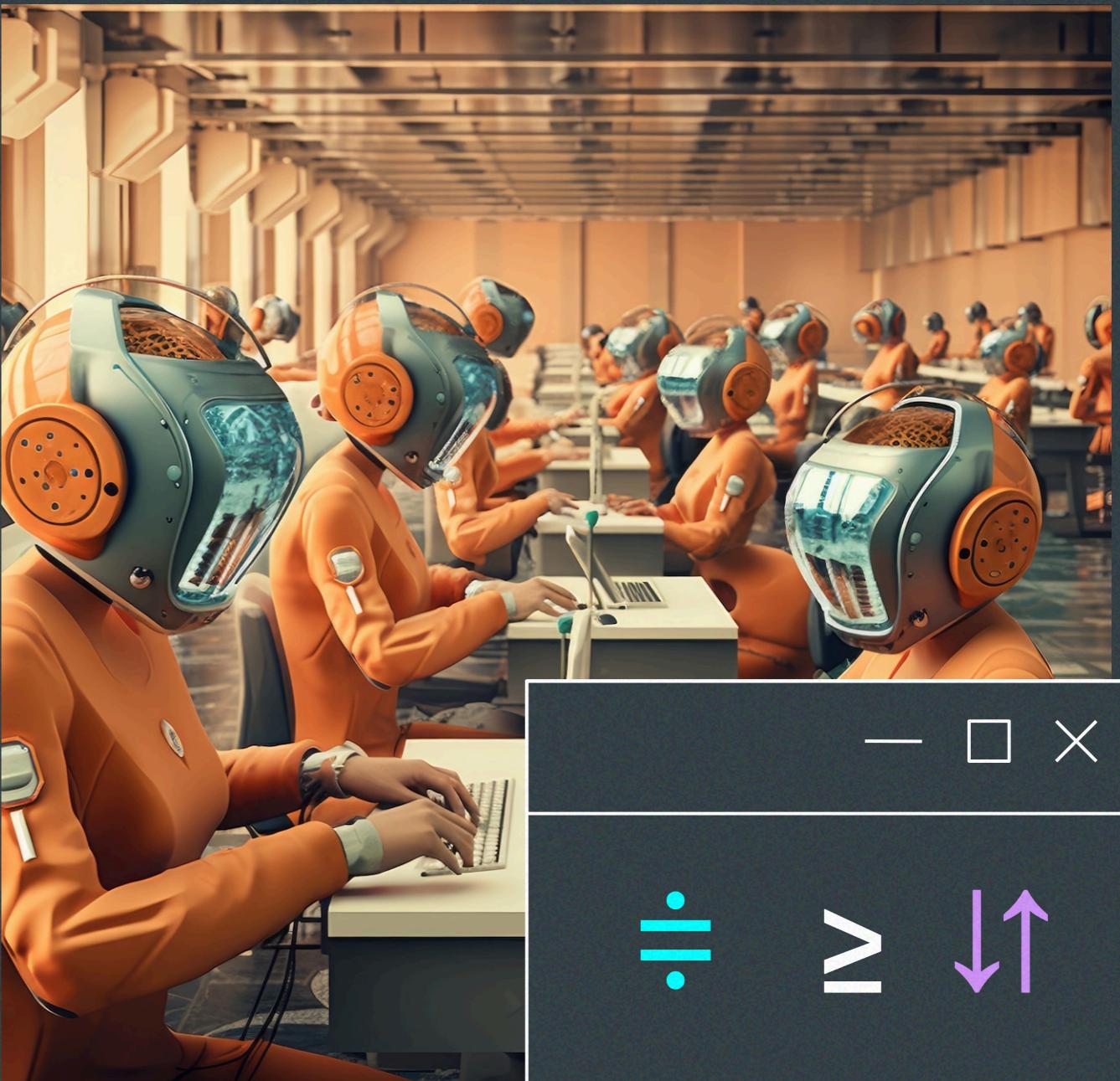
# Challenges

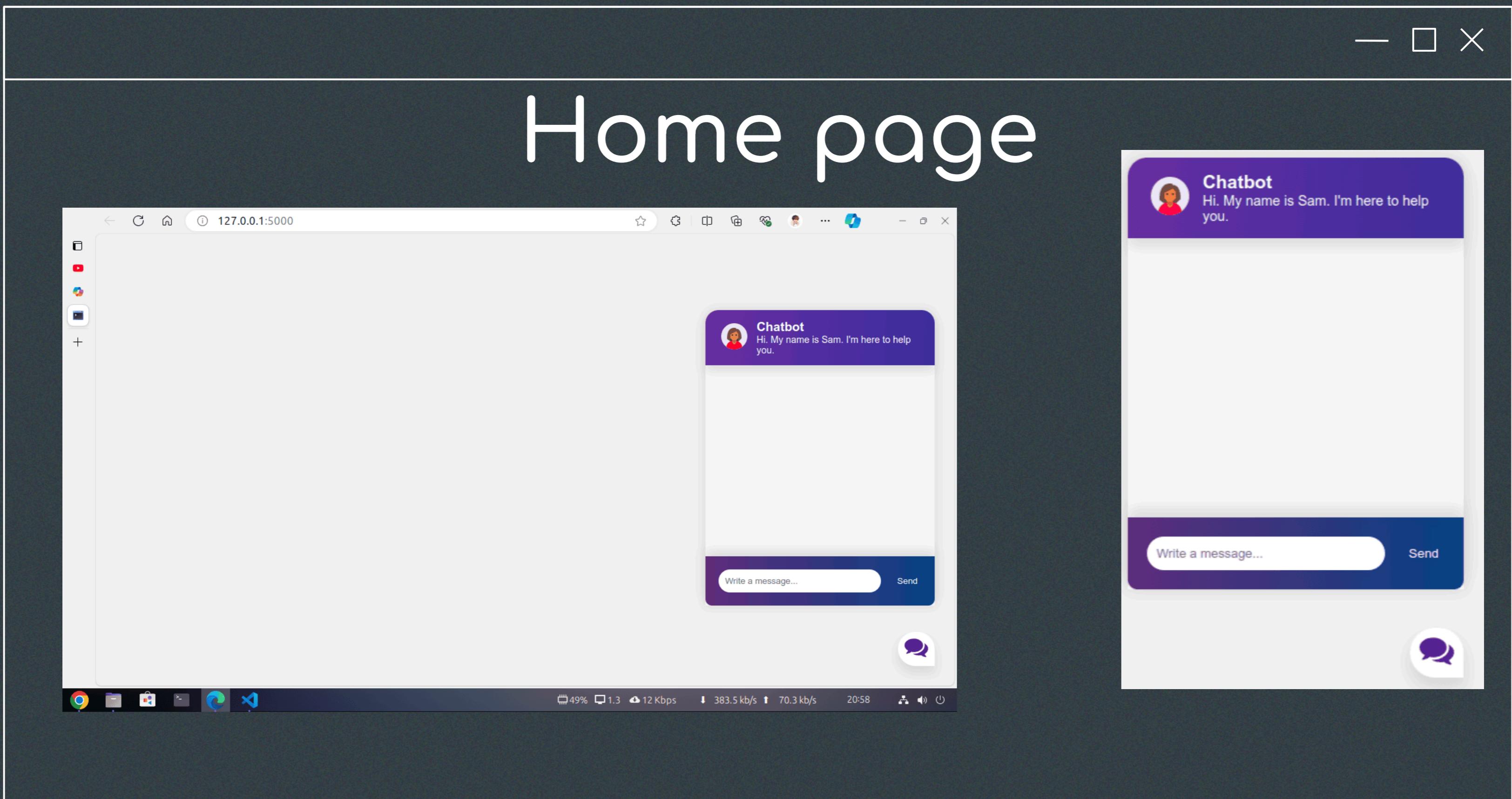
- Inaccurate Answer

chatbot might deliver responses that are off-topic or don't fully address the student's query. This can occur due to limitations in the LLM's ability to grasp the nuanced intent behind a student's question

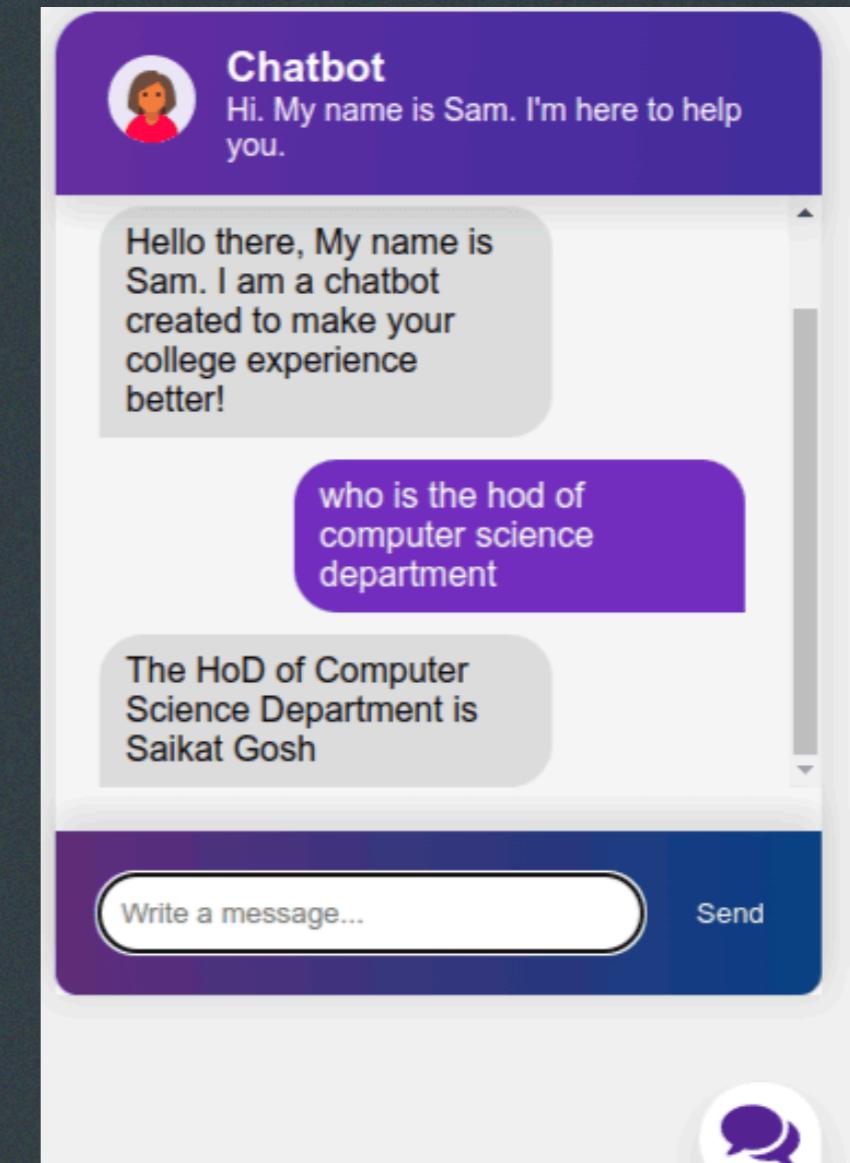
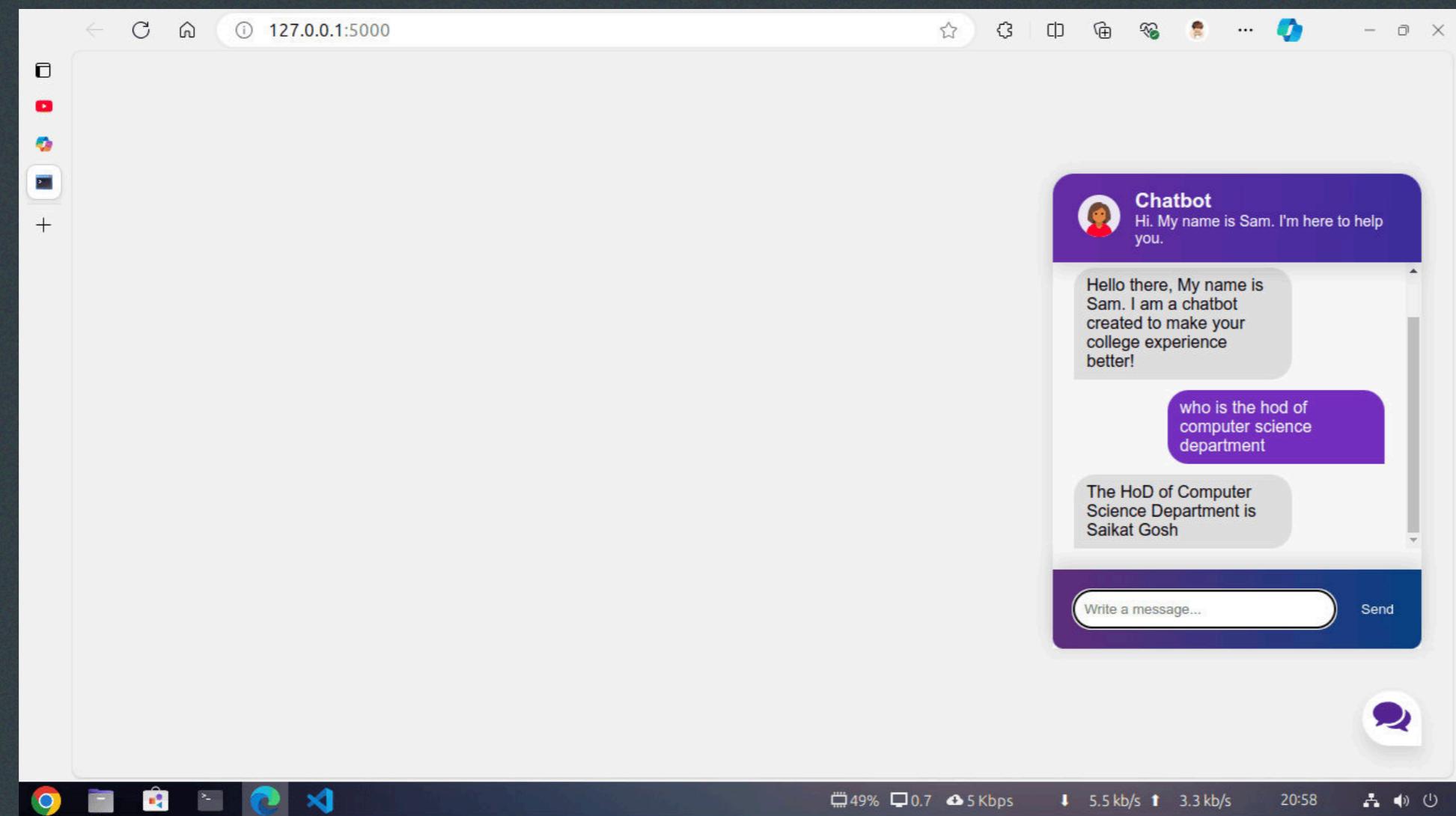
- Spelling Errors

The chatbot may struggle to comprehend student queries that contain typos or grammatical mistakes. This can be frustrating for users and hinder their ability to get the information they need.





# Chatting page



# Conclusion

In conclusion, this project has successfully developed a large language model (LLM) chatbot designed to serve as a comprehensive information resource for students of Balurghat College. The chatbot leverages natural language processing (NLP) techniques to understand student queries and provides direct answers on topics ranging from college policies to faculty expertise, extracurricular activities, and student life.

While the project has achieved significant progress, there is always room for improvement.

Thanks!

